**Odds & ORs:**

**PART 2: numerical explanatory variable**

***Solutions***

1. **Med School Acceptance: MCAT scores**   
   Open the **MedGPA** data set (from the Stat2Data package) in R and use ?MedGPA to learn about the data.

The variable “MCAT” contains the total MCAT scores for these 55 students. There is a lot of variability in MCAT scores, and there are very few students who have the same value. Instead of looking at the odds of success (acceptance) for each individual score value, we’ll group them, following the procedure described on p. 472- 474 (Example 9.12).

* 1. First, we divide the range of MCAT scores into intervals with roughly equal numbers of cases. I have made (what I think is) a judicious division below into 5 intervals (however, there is no “right” number of intervals in these cases!).

Fill in the rest of the table.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Group # | # Cases | Range of MCAT | Midpoint of range | Admitted | | Proportion admitted | Odds of admission | Odds Ratios |
| Yes | No |
| 1 | 11 | 18 – 32 | 25 | 2 | 9 | 2/11 | 2/9 |  |
| 3.6 |
| 2 | 9 | 33 – 34 | 33.5 | 4 | 5 | 4/9 | 4/5 |
| 1.875 |
| 3 | 10 | 35 – 36 | 35.5 | 6 | 4 | 6/10 | 6/4 |
| 1.067 |
| 4 | 13 | 37 – 39 | 38 | 8 | 5 | 8/13 | 8/5 |
| 3.125 |
| 5 | 12 | 40 – 48 | 44 | 10 | 2 | 10/12 | 10/2 |
|  |

* 1. Interpret the odds of admission for Group 2.

The odds of admission for those with a MCAT range from 33-34 is 0.8, which means we expect 8 people with MCAT scores of 33-34 will be accepted for every 10 people denied.

* 1. Interpret (in context!) the odds ratio for an increase in MCAT score from Group 3 to Group 4.

An increase in MCAT score from the 35 – 36 range to the 37 – 39 range is associated with 1.067 times higher odds of being admitted to med school.

* 1. Create three vectors from the table above: “midpoints”, that contains the range midpoints for each group; “props”, that contains the proportion of admission for each group; and “odds”, that contains the odds from each group. You can create these vectors easily using (for example)

midpoints <- c(25, 33.5, …)

props <- c(2/11, 4/9, …)  
odds <- c(2/9, 4/5, …)

* 1. Make three scatterplots:

(1) proportion of success vs. MCAT score (“midpoints”);

(2) odds vs. MCAT score;

(3) log(odds) vs. MCAT score.

Look carefully at the difference between these three plots, and make sure you understand what each is plotting. What form do we expect in plot (1)? What form do we expect in plot (3)?

For (1), we expect an S-curve. For (3), we expect a straight line.

It’s hard to tell the difference between the two plots in this case because we have so few data points – only 5 groups!

Chart, scatter chart

Description automatically generated

2. **ICU Survival: Age**Open the **ICU** data set (from the Stat2Data package) in R and use ?ICU to learn about the data.

1. Make a boxplot of survival status by age. Does there appear to be a relationship between these two variables?

Chart, box and whisker chart

Description automatically generated

It looks like those who did survive are more likely to be younger, but it’s hard to say if there’s a significant relationship.

1. I have made (what I think is) a judicious division of Age into intervals (however, there is no “right” number of intervals in these cases!).

Fill in the rest of the table.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Group # | # Cases | Range of Age | Midpoint of range | Survived | | Proportion survived | Odds of survival | Odds Ratios |
| Yes | No |
| 1 | 29 | 15 – 29 | 22 | 27 | 2 | 27/29 | 27/2 |  |
| 1.50 |
| 2 | 30 | 30 – 49 | 39.5 | 27 | 3 | 27/30 | 27/3 |
| 2.70 |
| 3 | 26 | 50 – 59 | 54.5 | 20 | 6 | 20/26 | 20/6 |
| 1.11 |
| 4 | 24 | 60 – 64 | 62 | 18 | 6 | 18/24 | 18/6 |
| 0.68 |
| 5 | 27 | 65 – 69 | 67 | 22 | 5 | 22/27 | 22/5 |
| 0.93 |
| 6 | 23 | 70 – 74 | 72 | 19 | 4 | 19/23 | 19/4 |
| 2.53 |
| 7 | 23 | 75 – 79 | 77 | 15 | 8 | 15/23 | 15/8 |
| 0.94 |
| 8 | 18 | 80 – 92 | 86 | 12 | 6 | 12/18 | 12/6 |
|  |

**Remember**: We will not have to create a table like this when actually fitting a model, but we *will* need to create one to check our linearity condition.

1. Interpret the odds of survival for Group 2.

The odds of survival for people between age 30 and 49 is 27/3. This means we expect 27 people of this age range to survive in the ICU for each 3 that die.

1. Interpret (in context!) the odds ratio for Group 3 compared to Group 2.

OR for Group2 to Group3 = (27/3)/(20/6) = 2.7

The odds of survival in the ICU for 30-49-year-olds is 2.7 times the odds of survival for 50-59-year-olds.

1. Interpret (in context!) the odds ratio for Group 6 compared to Group 2.

odds of survival for Group 6 = 19/4

odds of survival for Group 2 = 27/3

OR for Group6 to Group2 = (19/4)/(27/3) = 0.53

The odds of survival in the ICU for 70-74-year-olds is 53% the odds of survival for 30-49-year-olds.

1. Create three vectors from the table above: “midpoints”, that contains the range midpoints for each group; “props”, that contains the proportion of survival for each group; and “odds”, that contains the odds from each group.   
   Then make three scatterplots:

(1) proportion of survival vs. age (“midpoints”);

(2) odds vs. age;

(3) log(odds) vs. age.

Look carefully at the difference between these three plots, and make sure you understand what each is plotting. What form do we expect in plot (1)? What form do we expect in plot (3)?

Again, for (1), we expect an S-curve. For (3), we expect a straight line. We can kind of (?) see those forms here.

Chart, scatter chart

Description automatically generated